SIMS: Computation of a Smooth Invariant Molecular Surface

Yury N. Vorobjev and Jan Hermans
Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill,
North Carolina 27599-7260, USA

ABSTRACT SIMS, a new method of calculating a smooth invariant molecular dot surface, is presented. The SIMS method generates the smooth molecular surface by rolling two probe spheres. A solvent probe sphere is rolled over the molecule and produces a Richards-Connolly molecular surface (MS), which envelops the solvent-excluded volume of the molecule. In deep crevices, Connolly's method of calculating the MS has two deficiencies. First, it produces self-intersecting parts of the molecular surface, which must be removed to obtain the correct MS. Second, the correct MS is not smooth, i.e., the direction of the normal vector of the MS is not continuous, and some points of the MS are singular. We present an exact method for removing self-intersecting parts and smoothing the singular regions of the MS. The singular MS is smoothed by rolling a smoothing probe sphere over the inward side of the singular MS. The MS in the vicinity of singularities is replaced with the reentrant surface of the smoothing probe sphere. The smoothing method does not disturb the topology of a singular MS, and the smooth MS is a better approximation of the dielectric border between high dielectric solvent and the low dielectric molecular interior. The SIMS method generates a smooth molecular dot surface, which has a quasi-uniform dot distribution in two orthogonal directions on the molecular surface, which is invariant with molecular rotation and stable under changes in the molecular conformation, and which can be used in a variety of implicit methods of modeling solvent effects. The SIMS program is faster than the Connolly MS program, and in a matter of seconds generates a smooth dot MS of a 200-residue protein. The program is available from the authors on request (see http://femto.med.unc.edu/SIMS).

INTRODUCTION

In recent years a variety of methods have been developed to calculate surfaces of molecules (Richards, 1977; Connolly, 1983a,b, 1985a,b, 1993; Perrot et al., 1992; Eisenhaber and Argos, 1993; Eisenhaber et al., 1995; Varshney et al., 1994; Pascual-Ahuir et al., 1994; Zauhar and Morgan, 1988; Zauhar, 1995). A molecule is considered a collection of interlocking spheres, and three definitions of surface are used for molecular modeling as defined by Richards (1977). The van der Waals (VW) surface is the external surface of atoms, each represented by a spherical ball of van der Waals radius. The solvent-accessible (SA) surface is generated by the center of a solvent probe molecule (modeled as a rigid sphere of finite radius) when this rolls about the VW surface of the molecule; the SA surface is equivalent to the VW surface of the molecule with atoms of extended radius (sum of actual van der Waals radius and van der Waals radius of the solvent molecule). The molecular surface (MS) is generated by the inward-facing surface of the solvent molecule probe sphere, when it rolls about the VW surface of the molecule. The molecular surface consists of three types of "faces" (i.e., "contact," "saddle," and "concave reentrant,") where the probe touches the molecule atoms at one, two, or three points simultaneously (Connolly, 1983a,b). The mo-

lecular surface envelops a solvent-excluded volume of the solute molecule, and is a better basis for representing hydrophobic association phenomena in water solvent than are the SA or VW molecular surfaces (Jackson and Sternberg, 1993, 1994, 1995). It is assumed that the MS is a good approximation of a dielectric border between high-dielectric polar solvent and the low-dielectric interior of the solute molecule in continuum dielectric methods of self-consistent reaction field (Zauhar and Morgan, 1988; Zauhar, 1995; Rashin and Namboodiri, 1987; Rashin, 1990; Grant et al., 1990; Sharp and Honig, 1990; Juffer et al., 1991; Vorobjev et al., 1992). However, the SA molecular surface area, which is simple to calculate, is also used in some empirical methods of implicit treatment of solvent (Eisenberg et al., 1986; Ooi et al., 1987; Vila et al., 1991; Ponnuswamy, 1993; Juffer et al., 1995).

Calculation of molecular properties on the MS and integration of a function over the MS need a numerical representation of the MS as a manifold $S(s_i, n_i, \Delta s_i)$, where s_i, n_i, Δs_i are the coordinates, normal vector, and area of a small element of the MS. Some methods that generate a dot molecular surface provide a numerical presentation of the MS as a collection of dot coordinates (with assigned dot area) and outward normal vectors (Connolly, 1983a,b, 1993); whereas other methods (Varshney et al., 1994; Zauhar, 1995) provide a triangulated dot molecular surface. Dot triangulation makes it possible do a continuous interpolation of surface functions on the MS. Methods of implicit modeling of solvation effects (Juffer et al., 1995; Jackson and Sternberg, 1994, 1995) and, particularly, the boundary element method of solution of the Poisson equation for a molecule in a polar solvent (Rashin, 1990; Juffer et al., 1991; Vorobjev et al., 1992; Vorobjev and Scheraga, 1997;

Zauhar, 1995; Bharadwaj et al., 1995), need a good numerical presentation of the MS to produce stable and reliable results.

The strictly mathematically defined molecular surface of complex molecules, like proteins, is a complex 3-D surface with singularities, i.e., the direction of the normal vector is not continuous in the vicinity of singular points of the MS. The singularities often appear in deep crevices, where the inward surface of the rolling solvent probe sphere can intersect itself and produce a self-intersecting singularity. Cusps and holes are two types of self-intersecting singularities of the MS that have been recognized (Connolly, 1985b; Zauhar, 1995). To our knowledge, an accurate calculation of the MS with accurate representation of self-intersecting singularities of the MS has never been described.

It can be assumed that a MS with smoothed singularities is a better approximation of the dielectric border between high-dielectric solvent and a low-dielectric molecule (Zauhar, 1995). Zauhar (1995) suggested a method of solvent probe deformation that completely avoids the occurrence of self-intersecting singularities in the MS and constructs a smooth molecular surface. A major deficiency of Zauhar's method is that it produces a smooth molecular surface which, in terms of topology and atomic areas, differs unpredictably from the strict, singular MS.

To achieve accurate numerical representation of surface functions and surface functionals of the MS, the dot MS must satisfy several conditions: 1) maximum homogeneity of dot distribution; 2) adequate smoothing of the MS in the vicinity of a singular points; 3) stability under variation of the dot density; 4) independence of rotation of the molecule; 5) stability under change in the molecular conformation. Development of a method of generation of the dot MS, which satisfies these five conditions and gives a good numerical representation of the MS of a molecule, is highly desirable.

We found that several available programs, MSEED (Perrot et al., 1992; Vorobjev et al., 1992), Varshney's program (Varshney et al., 1994), and Connolly's MS program (Connolly, 1983c) provide a dot MS of poor quality. The fast MSEED method (Perrot et al., 1992) is not completely general, as it can not handle a free saddle surface, and the MSEED-DOT program (Vorobjev, Performance of the MSEED-DOT program, unpublished results) has problems with deep concave and saddle faces (it does not remove a self-intersecting surface), and frequently produces results that are unstable to small perturbations of atomic coordinates. Varshney's method (Varshney et al., 1994) produces a triangulated surface with a large fraction of small triangles (Vorobjev, Performance of the Connolly MS, MSEED and Varshney methods with the boundary element method, unpublished results), i.e., the distribution of points on the molecular surface is nonuniform, and more importantly, this method lacks a general procedure for handling self-intersecting surfaces. Connolly's MS program (Connolly, 1983c) is the most reliable, i.e., it never fails and includes a numerical procedure for removing the dots on self-intersect-

ing surface segments. Connolly's MS program does not have satisfactory homogeneity of dot distribution, and produces a noninvariant dot distribution and atomic surface areas, when the molecule is rotated in a fixed conformation (Besler et al., 1990; Merz, 1992; Vorobjev, Performance of the Connolly MS, MSEED and Varshney methods with the boundary element method, unpublished results). None of these methods produces a smooth molecular surface.

This paper describes a new method (SIMS, smooth invariant molecular surface) for generating a smooth molecular surface. The paper 1) revises Connolly's method of generation of the MS, 2) describes all types of self-intersecting singularities of the MS and an exact method for calculating the singular regions of the MS, 3) describes a method of calculation of an invariant molecular surface (i.e., invariant under molecular rotation and translation) with quasi-homogeneous dot distribution, and 4) develops a general method of smoothing of singularities of the MS and generation of a molecular surface with specified minimum radius of curvature. The SIMS method calculates a dot SMS of good numerical quality, which can be used in a variety of implicit continuum models of calculation of solvation effects (Juffer et al., 1995; Jackson and Sternberg, 1995), and for molecular electrostatic calculations with the boundary element method in dielectric continuum models (Vorobjev et al., 1992; Vorobjev and Scheraga, 1997; Zauhar, 1995).

GENERATION OF SURFACE ELEMENTS

The SIMS algorithm starts from atomic coordinates a_i, an atomic radii set r_i, and a solvent probe sphere radius r_p. The neighbor list of atoms around atom i inside the sphere of extended radius $r_{ie} = r_i + r_p$ is defined by a cubic lattice method (Connolly, 1983c; Eisenhaber et al., 1995). Contact triplets of atoms $i < j < k$, which can simultaneously be touched by the solvent probe sphere, are defined. The solvent probe positions p_{ijk} and vertex coordinates v_i, v_j, v_k (i.e., contact points with the solvent probe sphere for the contact triplet of atoms i, j, k) are calculated by Connolly's equations (Connolly, 1983a,b). All solvent probe positions p_{ijk} for atomic pair i, j are calculated, and pairs of starting and stopping positions p, p' of solvent probe rotation around interatomic axis ij are defined, as illustrated in Fig. 1. Starting and stopping probe positions p, p' are associated, respectively, with vertices v_i, v_i' and v_j, v_j' and spherical triangles $(v_iv_jv_k)$ and $(v_i'v_j'v_l)$ on the solvent probe sphere. These spherical triangles form reentrant concave (RC) faces of the MS. Rotation of the solvent probe sphere between respective starting and stopping positions produces a saddle face S, which is a fraction of a side surface of a concave cylinder generated by rotation of the arc (v_i, v_j) of the solvent probe sphere around the interatomic axis ij by an angle ϕ_s, as shown in Fig. 1. The faces are joined together at common boundary arcs; a planar view of the MS between four atoms i, j, k, l is presented in Fig. 2. The joint between two faces is smooth, in the sense that there is a well-defined
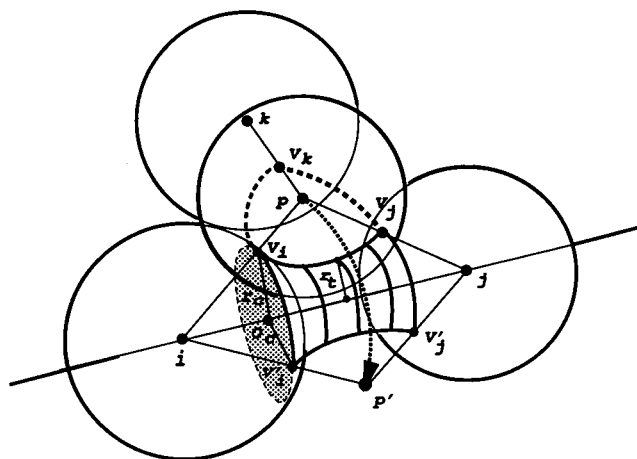
FIGURE 1   Generation of the molecular surface by a rolling solvent probe sphere. The contact points v_i, v_k, v_k between the probe sphere in position p and VW spheres of atoms i, j, k are the three vertices of a spherical triangle that produces a reentrant concave face. Rotation of the solvent probe sphere from position p to p' produces a saddle face (v_i, v_j, v_i', v_j'). The point O_s is the center of the atom-torus contact circle; r_c is a radius of the atom-torus contact circle; the saddle wrap angle is ($v_i O_c v_i'$); r_t is the radius of a slice of the saddle cylinder.

tangent plane at each point of the arc joining the faces, i.e., the direction of the normal vector of the tangent plane is a continuous function of the surface point. The contact (C) portion of the MS surface of atom i is a part of the VW spherical surface of the atom i, which results in a cutoff of the VW sphere by all atom-torus contact planes PL_{ij} between atom i and neighbor atoms j. The atom-torus contact plane PL_{ij} is orthogonal to the interatomic axis ij and passes through the contact circle (i.e., through the pair of vertices v_i, v_i'), as shown in Fig. 2. The areas of the reentrant
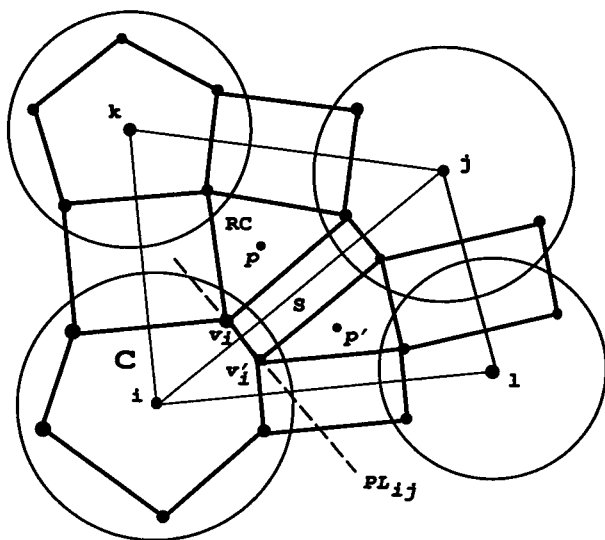

FIGURE 2   Illustration of planar mapping of the molecular surface. Formation of reentrant concave (RC), saddle (S), and contact (C) faces of the MS. i, j, k, l are atom centers; p, p' are solvent probe positions; PL_{ij} is a plane passing through the atom-torus contact circle.

concave, saddle, and contact convex faces, shown in Fig. 2, can be calculated analytically with the formulas in table 3 of Connolly (1983a).

Self-intersecting molecular surface

The simple molecular surface generated by Connolly's method (Connolly, 1983a,b), illustrated in Fig. 2, becomes complicated when the interatomic axis ij intersects the solvent probe sphere, as shown in Fig. 3. In this case, the MS becomes self-intersecting, i.e., a piece of the MS generated by the inward face of the solvent probe in a first position p can be overridden when the solvent probe moves to a second position p'. Self-intersection of the MS creates two problems. First, removal of the self-intersecting parts of the MS is necessary to obtain the correct MS, i.e., the surface that contains the solvent-excluded volume. Second, the correct MS in the vicinity of the self-intersecting regions is singular in the sense that the direction of the normal vector $n(s)$ is discontinuous at the point s, s being a singular point of the MS.

We describe an accurate method for defining and removing the self-intersecting parts of the MS and calculate the correct MS. Three types of self-intersecting MS are described. Removal of the normal vector discontinuity (i.e., smoothing of the MS) is important for improving descriptions of surface functions and surface functionals, and methods of self-consistent reaction field calculation (Rashin, 1990; Juffer et al., 1991; Vorobjev et al., 1992; Vorobjev and Scheraga, 1997; Sitkoff et al., 1994; Zauhar, 1995; Bharadwaj et al., 1995).

Free saddle cusps

If the solvent probe sphere can rotate around the interatomic axis ij between two contact atoms over the full angle of $\phi_s = 2\pi$, as shown in Fig. 3, and the solvent probe sphere arc ($v_i v_j$) intersects with the interatomic axis ij, then the saddle surface between atoms i and j converts into two
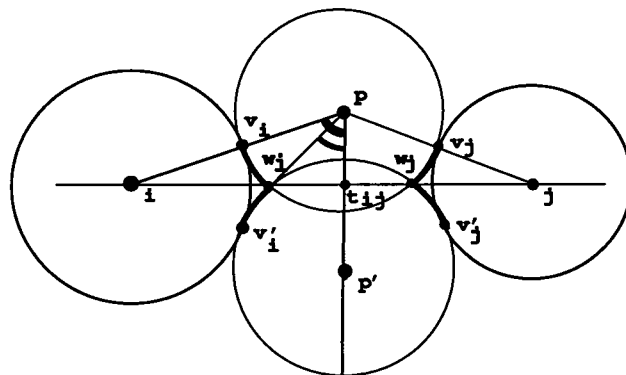

FIGURE 3   Formation of singular free saddle cusps (shown in bold). p, p' are twofold symmetrical positions of the solvent probe sphere against the interatomic axis ij. Angles $\theta_{max} = (v_i p t_{ij})$, $\theta_{min} = (v_i p w_i)$, respectively.

separate cone cusps. The cusps are generated by rotation about the axis ij of arcs $(v_i w_i)$ and $(v_j w_j)$, respectively, as shown in Fig. 3. The arcs $(v_i w_i)$, $(v_j w_j)$ can be found and the respective areas of a cusp can be calculated analytically as the surface of revolution of the respective arc by the general equation

$$\text{Area}_S = \phi_s r_p [h_{ij}(\theta_{max} - \theta_{min}) - r_p(\sin \theta_{max} - \sin \theta_{min})] \tag{1}$$

where $h_{ij} = |p - t_{ij}|$ is probe height, and the angles θ_{max} and θ_{min} are defined in Fig. 3. The point w_i is a singular point of the MS with undefined direction of the normal vector $\mathbf{n}(w_i)$; the same description applies to w_j.

Holes

The second case of a self-intersecting surface occurs when a contact triplet of atoms ijk has two symmetrical probe positions p and p' relative to the base plane $a_i a_j a_k$ passing through atoms i, j, k, as shown in Fig. 4. If the distance d'_{pp} between the probe positions p, p' is less than the probe diameter,

$$d_{pp'} = |p - p'| < 2r_p \tag{2}$$

then the two symmetrical reentrant concave faces, RC and RC', defined by the spherical triangles $(v_i v_j v_k)$ and $(v'_i v'_j v'_k)$, interpenetrate. Their self-intersection cuts off equal spherical cups in the RC and RC' faces and forms a closed-hole singular surface in these RC faces, i.e., a circle with center O (Fig. 4). The circle O contains the singular points w of the MS, where the direction of the normal vector $\mathbf{n}(w)$ is undefined. The new area of each RC face with a closed hole can be calculated analytically. The area of the spherical triangle $(v_i v_j v_k)$ (and symmetrical triangle $(v'_i v'_j v'_k)$) is decreased by the side area of the spherical cup sphere of height $\delta'_{pp} =$

FIGURE 4 Formation of a closed-hole singular surface. Atom centers are i, j, k; p and p' are probe positions related symmetrically via the atomic plane $a_i a_j a_k$; point O, the center of the closed hole, is the point of intersection between interprobe axis pp' and atomic plane $a_i a_j a_k$; point C is the point of intersection between interprobe axis pp' and the probe sphere p.

$|OC|$, where

$$\delta_{pp'} = \frac{d_{pp'}}{2} - r_p \tag{3}$$

The spherical cups are defined by the intersection between solvent probe spheres in the positions p and p' and the base plane $a_i a_j a_k$, as shown in Fig. 4. The side area of the spherical cup of the height $\delta pp'$ of the probe sphere is equal to

$$\Delta_{S_{RC\,hole}} = 2\pi r_p \delta pp' \tag{4}$$

A hole of more complicated shape, an open hole, occurs for deep, self-intersecting symmetrical reentrant concave faces when arc $(v_i v_j)$ (or one of arcs $(v_i v_k)$, $(v_j v_k)$) intersects with the base plane $a_i a_j a_k$. For example, if arc $(v_i v_j)$, which lies in the plane $a_i a_j p$, intersects the interatomic axis ij, and, consequently (cf. Fig. 4), the interatomic axis ij intersects the circle O, then the circle O will be opened and a singularity of open-hole type will be formed. Analytical calculation of the area of the self-intersecting RC, RC' faces forming an open hole is very complicated, and it is hardly necessary, because the case of deep, self-intersecting symmetrical RC faces is expected to be very rare for an arbitrary molecule. Therefore we use a numerical estimate of the area of deep self-intersecting symmetrical RC faces.

In general, for any type of hole, a surface dot s associated with area Δs on the RC face, generated by the solvent probe position p, survives the intersection with the symmetrical RC' face generated by the solvent probe position p', if the point s is above the atomic plane $a_i a_j a_k$,

$$(p - s) \cdot pp' < \frac{1}{2} |pp'|^2 \tag{5}$$

By placing dots on the RC faces and counting the area of surviving dots, the area of two self-intersecting RC faces with open holes can be estimated.

Concave edges

A nonsymmetrical general case of a self-intersecting molecular surface occurs for deep saddle faces, when the solvent probe sphere rotates between atoms i and j, and intersects the interatomic axis ij, as shown in Fig. 5. If p and p' are the starting and stopping positions of the probe rotation around the interatomic axis ij (Fig. 5), then the surviving inward surface of the rotating solvent probe, which is outside the solvent-excluded volume, is defined by the intersection between the probe spheres in the positions p and p'. The surviving reentrant MS formed by the inward surface of the solvent probe sphere consists of four surface elements, as shown in Figs. 5 and 6. The first one is a piece of the saddle face, which is a side surface of the cone $(v_i w_i v'_i)$ generated by the arc $(v_i w_i)$ while the probe rotates from position p to position p'. The second surface element is the RC face, the spherical triangle $(v_i v_j v_k)$; the third
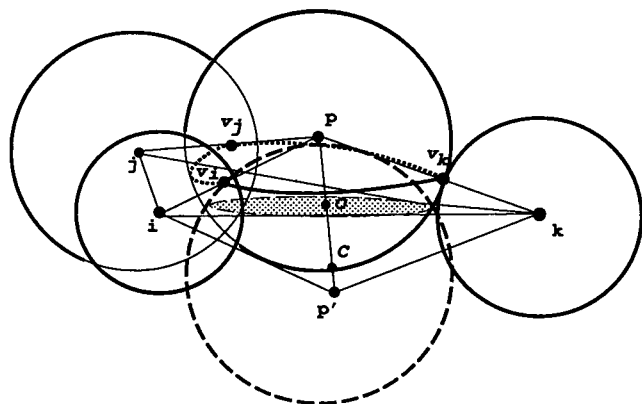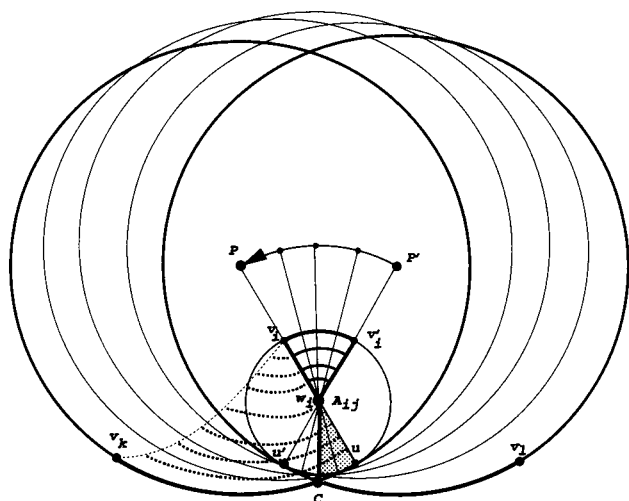
FIGURE 5 Formation of a concave-edge singular surface. The solvent probe sphere rolls between positions p and p'; these spheres are drawn with heavier lines, and intermediate positions are shown with thin lines. Point A_{ij} is a projection of the interatomic axis between atoms i, j; w_i is a point of intersection of arcs (v_iv_j), $(v_i'v_j')$ with the interatomic axis ij. The shaded area is part of the spherical triangle v_i, v_j, v_k, which lies inside the solvent probe sphere p'.

surface element is the RC$'$ face, the spherical triangle $(v_iv_jv_l)$. The fourth element is a piece of the saddle face, i.e., the side surface of the cone $(v_jw_jv_j')$, which is generated by the arc (v_jw_j) when the probe is rotated from position p to position p'. The solvent probe spheres in positions p and p' intersect over a small circle C with radius r_c and center O_c; therefore a part of the spherical triangle $(v_iv_jv_k)$ (*shaded area* in Figs. 5 and 6) lies inside the solvent probe sphere p', and a part of the spherical triangle $(v_i'v_j'v_l)$ lies inside the solvent probe sphere p. This self-intersecting removable part ΔRC of the spherical triangle $(v_iv_jv_k)$ is the solvent probe sphere surface between two arcs, an arc (w_i, u, w_j) of great circle of radius of r_p and an arc (w_i, C, w_j) of small circle of radius r_c, shown as shaded areas in Fig. 6 and in Fig. 5. Area ΔA_{RC} of the deleted portion of the surface of

the spherical triangle $(v_iv_jv_k)$ can be calculated analytically by applying the Gauss-Bonnet formula (Connolly, 1983a; Do Carmo, 1976), which defines area A of a polygon on a sphere of radius r_p:

$$\sum_{vertex} (\pi - \alpha_v) + \sum_{edges} \int k_e \, dl + \frac{A}{r_p^2} = 2\pi \qquad (6)$$

where α_v is an angle at vertex v, and k_e is the geodesic curvature of edge e. Using the value of geodesic curvature of arcs of the small circle r_c, $k_e = d_c/(r_pr_c)$ (Connolly, 1983a), the area between two arcs is given by

$$\frac{\Delta A_{RC}}{r_p^2} = \alpha + \beta - \frac{d_c}{r_p}\phi_c \qquad (7)$$

where α, β are the angles between arc (w_i, u, w_j) of great circle radius r_p and arc (w_i, C, w_j) of small circle radius r_c at vertices w_i, w_j, respectively; d_c is the displacement of center of the small circle O_c from the center p of the solvent probe sphere $2d_c = |p' - p|$; the radius of the small circle is $r_c^2 = r_p^2 - d_c^2$; the arc length of the small circle is $\phi = $ arccos$(2h_c/r_c)$, where h_c is the distance between center O_c of the small circle and interatomic axis ij; the position of the center of the small circle is $O_c = \frac{1}{2}(p + p')$.

GENERATION OF DOT MOLECULAR SURFACE

In the Introduction, we defined five conditions of good numerical representation of the molecular surface. The SIMS method has been developed to satisfy these conditions as much as possible, while avoiding dot MS triangulation to achieve good performance. In this section we describe the generation of points on a spherical template and methods of dot generation on reentrant concave, saddle, and contact faces of the MS.

Dots on a sphere

Generation of a homogeneous distribution of points on a sphere is not a trivial problem, several methods for which of have been discussed in the literature (Le Grand and Merz, 1993; Eisenhaber et al., 1995; Bliznyuk and Gready, 1996; Spackman, 1996). Random point selection (Woods et al., 1990) is nonreproducible unless a large number of points are used, $\sim$1500 per atom, which is unacceptably large for macromolecules. A constant arc length algorithm (Connolly, 1983c) generates latitude and longitude arcs of approximately constant length $l \simeq 1/\sqrt{d_s}$, where d_s is the density of points per Å^2, between points, and produces a quasi-uniform dot distribution of a specified density on a sphere. Because of the discrete nature of the problem, the allowed number of points is a stepwise function, with granularity, $N_p = 6, 12, 14, 20, 23, 30, 34, 44, 49, 60, 64, 80, 85, 100, 106, 126, 129$, etc. (Spackman, 1996).

Geodesic templates obtained by tessellation of the triangular faces of the regular icosahedron (Pascual-Ahuir et al.,
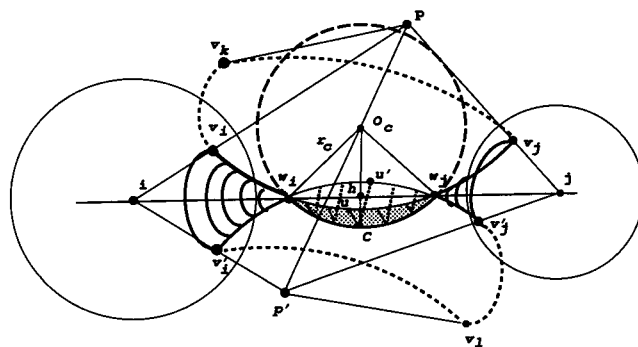

FIGURE 6 Formation of a concave-edge singular surface. Designations are the same as in Fig. 5. O_c is the center of the contact circle between solvent probe spheres p, p'; r_c is a radius of the contact circle. The shaded area is part of the spherical triangle v_i, v_j, v_k, which lies inside of the solvent probe sphere p'.

1994; Eisenhaber et al., 1995) produce approximately isotropic arrays of points on the surface of a sphere. The isotropicity of these geodesic templates can be further refined, by minimization of a Q potential (Bliznyuk and Gready, 1996), which is equivalent to the Coulombic energy of interaction between points of equal charge on the sphere, or a U potential, which maximizes the arc distance between points on the sphere (Le Grand and Merz, 1993). The granularity of the geodesic templates, i.e., the allowed number of points N_p on a sphere, is equal to 12, 20, 32, 42, 60, 80, 92, 122, 162, 256, etc. (Spackman, 1996), and is quite sparse compared to the granularity of the constant arc length algorithm (Connolly, 1983c). In some applications, like self-consistent reaction field calculations (Rashin, 1990; Juffer et al., 1991; Vorobjev et al., 1992; Vorobjev and Scheraga, 1997; Zauhar, 1995), a smooth granularity is important for achieving fast calculations without loss in accuracy.

The SIMS method uses the simple and flexible constant arc-length algorithm (Connolly, 1983c) to generate surface point templates on the solvent probe sphere and van der Waals spheres of atoms. The SIMS algorithm calculates a variable area Δs_i assigned to dot i, which is considered a spherical rectangle with vertices at $\theta_i \pm \frac{1}{2}\delta\theta_i$, $\phi_i \pm \frac{1}{2}\delta\phi_i(\theta_i)$, where $\delta\theta_i = \pi/n_{gcr}$, $\delta\phi_i = 2\pi/n_{alt}$, and where n_{gcr}, n_{alt} are the number of dots along a great circle and a small circle at given altitude θ_i. The variable-size spherical rectangles exactly cover the surface of a sphere, and this method improves the stability of calculations of surface-related properties compared to Connolly's algorithm, which assigns an equal area to each dot (Connolly, 1983c). Analysis shows that the standard deviation of the variable dot area Δs_i is in the range of 11%, the greatest uniformity of the dot distribution being observed in the equatorial region of the sphere.

Dots on a reentrant concave face

To distribute dots over a spherical triangle, we use the precalculated dot template of the solvent probe sphere. The solvent probe sphere dot template is placed in the respective probe position p, and template dots are projected on the spherical triangle $(v_iv_jv_k)$ (Fig. 1). To achieve independence of the projected dots on rotation of the molecule, the dot template is calculated in the local coordinate system of the spherical triangle. The local coordinate system is defined by the three vectors from the center p of the solvent probe sphere to the vertices v_i, v_j, v_k. The longest median arc of the spherical triangle is oriented to coincide with the equatorial plane of the probe sphere, and the first dot is placed at the center of the triangle. The area of the spherical triangle is calculated analytically by the Gauss-Bonnet formula (6), and the areas of projected dots are normalized to the analytical area. It should be mentioned that Connolly's MS program (Connolly, 1983c) projects dots onto a spherical triangle from a dot template in fixed orientation relative to a laboratory coordinate system. Therefore the resulting dot

distribution depends on the orientation of the spherical triangle, i.e., the local coordinate system of the spherical triangle, with respect to the fixed laboratory coordinate system, and consequently changes when the molecule rotates.

Dots on a reentrant saddle face

The dot distribution on saddle faces is generated by using a method of equal arc length in two orthogonal directions along two lines of main curvature of the surface. The first one is along the solvent probe sphere arc (v_iv_j), and the second one is along the torus circle of rotation around the interatomic axis ij (Fig. 1). The polar angle $\delta\phi_t$ of rotation along a torus circle between two neighbor dots is equal to

$$\delta\phi_t(r_t) = \delta\theta_p \frac{r_p}{r_t} \tag{8}$$

where $\delta\theta_p$ is the polar angle between these dots along the large circle of the solvent probe, and r_t is the average radius of the slice of the saddle cylinder for the dot. The surface elements on the saddle face have a rectangular shape, and the area assigned to the dot is calculated analytically by the general formula (Eq. 1). This approach allows one to obtain a quasi-uniform dot distribution on saddle faces.

Straightforward mapping of a saddle face to a planar rectangle, as used by others (Connolly, 1983a,c; Zauhar, 1995), does not produce a quasi-uniform dot distribution on a saddle face, because in 3-D space distances between dots along the solvent probe sphere arc (v_iv_j) and in the orthogonal direction on the surface along the torus circle can be substantially different (Fig. 1).

Dots on a contact face

The dot distribution on a contact face of atom i is obtained in three steps: 1) a dot distribution template on the atomic sphere of given radius is calculated; 2) the dot template is placed in position a_i of the atom i in an orientation related to the local coordinate system, which is defined by two nearest-neighbor atoms j, k; 3) dots that lie inside torus ij, i.e., on the spherical cup defined by the atom-torus contact plane, as shown in Fig. 2, are removed. The resulting polygon on the VW sphere of atom i has a quasi-uniform distribution of dots. The total area of the spherical polygon can be calculated by the Gauss-Bonnet formula (Eq. 6). The areas of the dots of the contact face can be normalized to the analytical area of the spherical polygon.

SMOOTHING OF SINGULARITIES OF THE MS

The SIMS method uses a new method, a smoothing probe, to smooth singular regions of the MS. This method replaces a singular portion of the MS by the smooth outward surface of a smoothing probe sphere, which is rolled over the inside of the MS. As has been shown above, three types of self-intersecting singular regions of the MS (closed hole, open
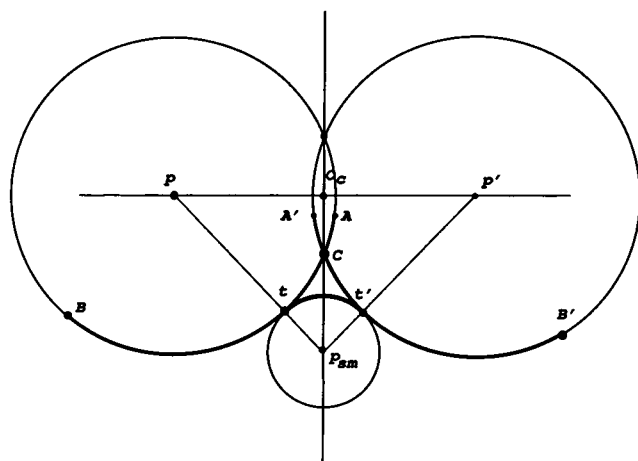
FIGURE 7   Smoothing of a singular molecular surface by rolling a smoothing probe sphere. p, p' are the solvent probe spheres; p_{sm} is the smoothing probe sphere. BCB' is a singular molecular surface; (t, C, t') is a singular portion of the MS in a vicinity of the singular point C, which is replaced by the smooth surface (t, t') of the smoothing probe sphere.

hole, and concave edge) are produced by intersection of the inward surface of two solvent probe spheres in positions p, p' (Figs. 3–6). Suppose that a singular portion of the MS is produced by intersecting solvent probe spheres in positions p, p', as shown in Fig. 7. The arcs (AC) and $(A'C)$ represent the self-intersecting removable pieces of the surface, and the arcs (CB) and (CB') are the actual singular molecular surface. It can be seen that the MS in the vicinity of the singular point C can be smoothed by rolling a smoothing probe sphere of radius r_{sm} over the outward surface of the intersecting solvent probe spheres p, p' around interprobe axis pp', as shown in Fig. 7. The MS generated by the solvent probe arcs C, t and C, t' is replaced by the surface generated

by the arc t, t' of the smoothing probe sphere in the position p_{sm}, where the points t, t' are the tangent points between the smoothing probe sphere and the solvent probe sphere in the positions p, p', respectively. The dots s of the MS are removed and replaced by the dot template s' on the arc of the smoothing probe sphere. The starting and stopping positions of the rotation of the smoothing probe sphere around the interprobe axis pp' are defined as the points w_i, w_j (Fig. 6). The radius of the smoothing probe is equal to zero at the points w_i, w_j and linearly grows to the regular value r_{sm} when the smoothing probe rolls away from the points w_i, w_j. For the closed hole, the rotation angle of the smoothing probe sphere is equal to 2π, and the probe's radius is constant. Singular cusps of free saddles (Fig. 3) are smoothed by replacing the singular side surface of the cone $(v_i w_i v_j)$ by the spherical cup t, t' of the inscribed smoothing probe sphere, as illustrated in Fig. 7. The minimum radius of curvature of the smoothed MS (SMS) is equal to the radius of the smoothing probe sphere r_{sm}. The smoothing probe sphere must have smaller radius than the solvent probe and any atom of the molecule. For practical calculations of the SMS, a reasonable value of r_{sm} is ≤ 0.5 Å.

RESULTS AND DISCUSSION

Table 1 shows the number of dots N_C, N_S, N_{RC} on the contact, saddle, and reentrant concave faces of the SMS of three molecules, i.e., a 17-residue peptide in the α-helical conformation, the 64-residue protein eglin, and the 224-residue immunoglobulin G1, calculated by the SIMS program. It can be seen that more than two-thirds of the total area of the MS belongs to saddle or reentrant concave faces, which provides an argument in favor of accurately representing these types of faces of the MS. The average areas

TABLE 1   Areas and number of dots for contact, saddle, and reentrant concave faces

Molecule	Residues	Atoms	S_C* N_C# Δs_C§ $\delta \Delta s_C$¶	S_S N_S Δs_S $\delta \Delta s_S$	S_{RC} N_{RC} Δs_{RC} $\delta \Delta s_{RC}$
17-Residue peptide‖	17	278	492	540	372
			2123	3361	1740
			0.23	0.16	0.21
			0.11	0.25	0.21
Eglin	64	1029	1210	1352	834
			4422	6446	3577
			0.27	0.21	0.23
			0.15	0.29	0.25
Immunoglobulin	224	2161	2940	3428	2653
			10742	16569	11010
			0.27	0.21	0.24
			0.19	0.34	0.27

Areas in Å^2.

* S_C, S_S, S_{RC} are the total areas of contact, saddle and reentrant concave faces, respectively.

N_C, etc. are the total numbers of dots on these faces.

§ Δs_C, etc. are the average areas per dot for these faces.

¶ $\delta \Delta s_C$, etc. are the standard deviations of dot areas for these faces, in %.

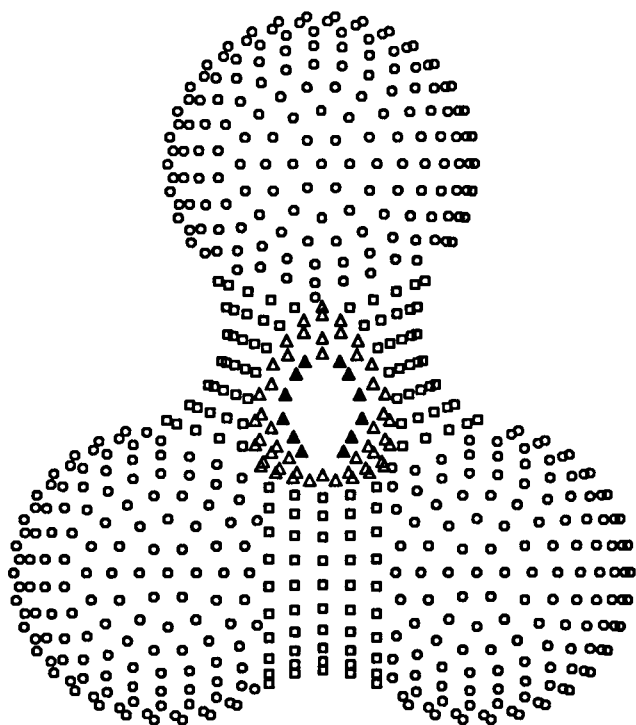‖ 17-Residue peptide Ace-ETGTKAELLAKYEATHK-NMe in the α-helical conformation.

uniformity of the dot distribution of the generated molecular surfaces. Visual evidence of the quasi-uniformity of the dot distribution on the smoothed MS is shown in Fig. 8, which depicts the dot distribution on the SMS of a three-atom system. The MS of this three-atom system has a singularity of closed-hole type on the reentrant concave faces. Smoothing the MS by the smoothing probe method removes dots in the vicinity of the singularity and adds dots on the surface of the spherical smoothing probe, as described above. Table 2 shows the number (N_{sing}) of singular concave edges found in the MS of three proteins. One sees that this number is quite large for peptide and protein molecules, which suggests that an accurate description of self-intersecting singular regions of the MS and correct removal of overridden segments of the molecular surface are important. The new smoothing method is a well-controlled procedure that does not disturb the topology and insignificantly disturbs the atomic areas of the singular, strict MS.

It can be seen from Table 2 that the CPU time for the SIMS method is in the range of seconds for a protein with several thousand atoms; calculation by SIMS is twice as fast as Connolly's MS program (Connolly, 1983c). The SIMS computation scales as $O(N_{at})$.

It should be noted that Connolly's MS program generates a numerical presentation of the dot molecular surface of a molecule, i.e., the number of dots, dot positions, and total area, which are not invariant during rotation of the molecule (Besler et al., 1990; Merz, 1992). It is shown in Table 2 that the standard deviation ΔS_{rot} of the area of the MS of a randomly rotated molecule is in the range of 0.3–1.0% of the total area. This deficiency of Connolly's program is due to the method of projection of dots on contact and reentrant faces from a spherical dot template that is kept in a fixed laboratory coordinate system. The SIMS method uses a local coordinate system for each contact triplet of atoms, which makes the result dot distribution and the MS surface area invariant to the molecular orientation.

Figs. 9 and 10 demonstrate stability of the SIMS numerical representation of the molecular surface versus dot density. The areas were calculated by Connolly's MS program (Connolly, 1983c) and by the SIMS program. It can be seen that the SIMS method produced much more stable results than the MS program and converges well when the dot size is decreased. The SIMS results show insignificant depen-

FIGURE 8   Smooth molecular dot surface for a three-atom system. Radii of the atoms and solvent probe are 2.0 and 1.4 Å, respectively. ○, Dots on contact faces; □, dots on saddle faces; △, dots on reentrant concave faces; ▲, smoothing probe dots, which are replacement dots in the vicinity of a singular circular line of a closed-hole singular region.

per dot for the different types of faces are within 40%. The standard deviation does not exceed 19% for the contact faces. The lowest value of $\delta\Delta s_C$ is found for the immunoglobulin, with CH_n groups treated as united atoms. The 17-residue peptide with all hydrogen atoms has the largest values of $\delta\Delta s_C$, $\delta\Delta s_S$, $\delta\Delta s_{RC}$. A careful investigation (Bliznyuk and Gready, 1996) of the uniformity of dot distributions on the unit sphere, generated by different methods of tessellation, reports standard deviations of dot areas in the range of 7–11%. The values of $\delta\Delta s_S$, $\delta\Delta s_{RC}$ for saddle and reentrant concave faces are twice as large as the value $\delta\Delta s_C$, because generation of dots on these faces is more complicated and is subject to more geometrical restrictions than the generation of dots on a sphere. These reasonably low values of the fluctuation of areas per dot confirms the quasi-

TABLE 2   Timing of the SIMS program

Molecule	Atoms	t_{SIMS}*	t_{MS}[#]	N_{sing}[§]	S_{MS}[¶]	ΔS_{rot}[‖]	S_{SIMS}***
17-Residue peptide	278	2.0	3.5	43	1409.3	9.1	1403.7
Eglin	1029	3.1	5.1	64	3412.8	12.3	3395.8
Immunoglobulin	2161	6.5	17.2	176	9027.9	19.7	9021.7

* CPU time in s (SGI Power Onyx R10000 processor).
[#] CPU time for Connolly's MS program (Connolly, 1983c).
[§] Number of singular regions (i.e., holes, concave edges).
[¶] Average total area of the MS over 100 random rotations of the molecule calculated by Connolly's MS program (area in Å²).
[‖] Standard deviation for the total area of the MS over 100 random rotations of the molecule for Connolly's MS program.
*** Total area of the SMS calculated by the SIMS program (area in Å²).
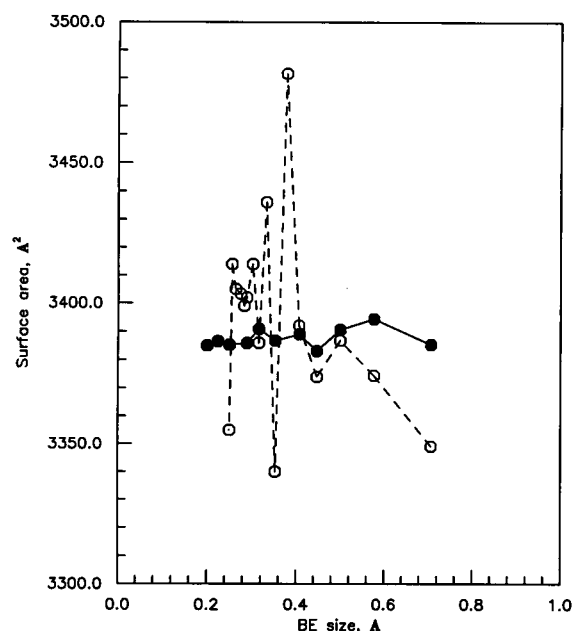
FIGURE 9   Area of the molecular surface of eglin as a function of the average dot size. ●, Results of the SIMS method; ○, results of Connolly's MS program (Connolly, 1983c).

FIGURE 11   Total area of the SIMS molecular surface of the protein eglin along a molecular dynamics trajectory. The trajectory was calculated with the SigmaX program (Hermans, 1995), with the cedar force field, explicit SPC water molecules, periodic boundary conditions, an integration time step of 2 fs, the Shake method to constrain bond lengths, and a cutoff radius of 10 Å for nonbonded interactions.

dence on the dot density, because the total area is calculated as a sum of areas of all dots of the SMS, where the area of each dot is calculated analytically, whereas in the MS program the dot areas on a dot template sphere (the solvent probe or atomic spheres) are all equal, and the number of surviving dots in singular regions of the MS is a step function. Fig. 10 shows the free energy of polarization in a high-dielectric solvent based on calculation of the polarization charge density on the molecular surface of the protein eglin. The polarization charge densities have been calcu-

lated by the multigrid boundary element method (Vorobjev and Scheraga, 1997) with the dot MS generated by the Connolly's program and the SIMS program. This test demonstrates a significant superiority of the SIMS method over the Connolly's MS program in its ability to calculate a numerically stable dot MS.

To estimate the stability of the SIMS method under small molecular conformational changes, we have calculated the SMS area and free energy of polarization of water solvent along a 1-ps molecular dynamics trajectory of the protein eglin for a sequence of conformations at intervals of 10 fs (or five time steps). It can be seen that the total molecular surface area (Fig. 11) and partial atomic surface areas (Fig. 12) are smoothly varying functions. Fig. 13 shows the free
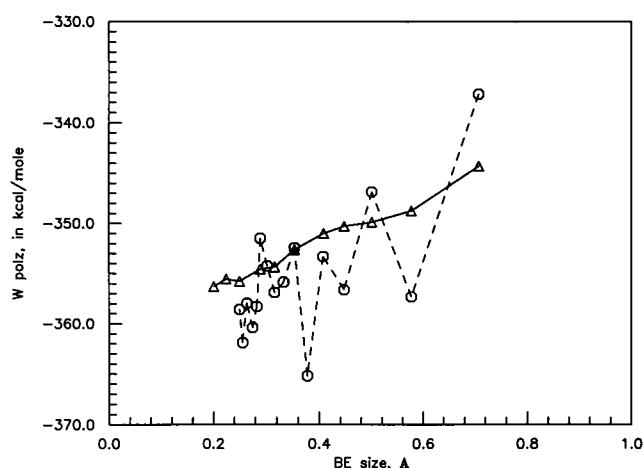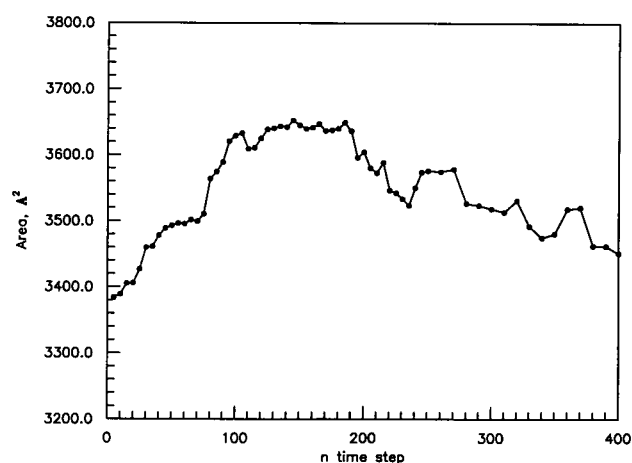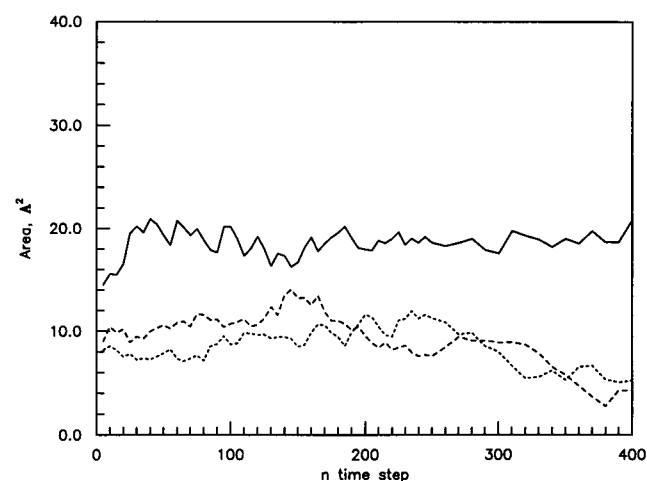

FIGURE 10   Free energy of polarization of water solvent by protein eglin, calculated by the multigrid boundary-element method (Vorobjev and Scheraga, 1997), as a function of the dot size used to calculate the dot MS. △, Results from the SIMS method; ○, results from Connolly's MS program (Connolly, 1983c).

FIGURE 12   Exposed areas of three atoms of eglin along a molecular dynamics trajectory: N^ζ of Lys 2 (———), O of Ser 3 (·····), and C^α of Glu 17 (— — —).
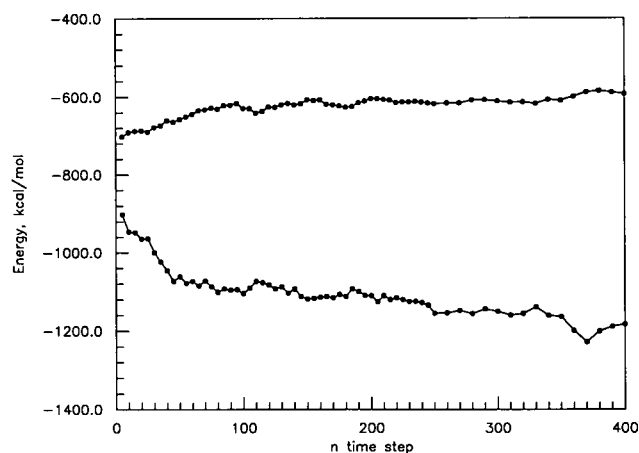
FIGURE 13   Free energy of polarization of water solvent (●) and intra-protein Coulombic energy (○) of eglin along a molecular dynamics trajectory.

energy of polarization of water solvent and the intraprotein Coulombic energy of eglin. The free energy of polarization is a complex functional of the MS and varies as smoothly as the intraprotein Coulombic energy. The data presented in Figs. 11–13 are clear evidence that the SIMS method provides a representation of the molecular surface that is numerically stable versus small conformational changes of the molecule.

The SIMS method is implemented in fortran77 and is available from the authors on request (see http://femto.med.unc.edu/SIMS).

REFERENCES

Besler, B. H., K. M. Merz, Jr., and P. A. Kollman. 1990. Atomic charges derived from semiempirical methods. *J. Comp. Chem.* 11:431–439.

Bharadwaj, K., A. Windemuth, S. Sridharan, B. Honig, and A. Nicholls. 1995. The fast multipole boundary method for molecular electrostatics: an optimal approach for large systems. *J. Comp. Chem.* 16:898–913.

Bliznyuk, A. A., and J. E. Gready. 1996. Numerical calculation of molecular surface area. I. Assessment of errors. *J. Comp. Chem.* 17:962–969.

Connolly, M. L. 1983a. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16:548–558.

Connolly, M. L. 1983b. Solvent-accessible surfaces of proteins and nucleic acids. *Science.* 221:709–713.

Connolly, M. L. 1983c. QCPE Program no. 429. Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN.

Connolly, M. L. 1985a. Molecular surface triangulation. *J. Appl. Crystallogr.* 18:499–505.

Connolly, M. L. 1985b. Computation of molecular volume. *J. Am. Chem. Soc.* 107:1118–1124.

Connolly, M. L. 1993. The molecular surface package. *J. Mol. Graph.* 11:139–141.

Do Carmo, M. P. 1976. Differential Geometry of Curves and Surfaces. Prentice-Hall, Englewood Cliffs, NJ. 264.

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203.

Eisenhaber, F., and P. Argos. 1993. Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency. *J. Comp. Chem.* 14:1272–1280.

Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. 1995. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and dot surface contouring of molecular assemblies. *J. Comp. Chem.* 16:273–284.

Grant, J. A., R. L. Williams, and H. A. Scheraga. 1990. Ab initio self-consistent field, and potential-dependent partial equalization of orbital electronegativity calculations of hydration properties of N-acetyl-N'-methyl-alanineamide. *Biopolymers.* 30:929–949.

Hermans, J. 1995. Sigma Documentation. Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC.

Jackson, R. M., and M. J. E. Sternberg. 1993. Protein surface area defined. *Nature.* 366:638–638.

Jackson, R. M., and M. J. E. Sternberg. 1994. Application of scaled particle theory to model the hydrophobic effect: implication for molecular association and protein stability. *Protein Eng.* 7:371–383.

Jackson, R. M., and M. J. E. Sternberg. 1995. A continuum model for protein-protein interactions: application to the docking problem. *J. Mol. Biol.* 250:258–275.

Jawson, M. A., and G. T. Symm. 1977. Integral Equation Methods in Potential theory and Elastostatics. Academic Press, New York.

Juffer, A. H., E. F. F. Botta, A. M. Bert, B. A. M. van Keulen, A. Van der Ploeg, and H. J. C. Berendsen. 1991. The electric potential of macromolecule in solvent: a fundamental approach. *J. Comp. Phys.* 97: 144–171.

Juffer, A. H., F. Eisenhaber, S. J. Hubbard, D. Walter, and P. Argos. 1995. Comparison of atomic solvation parametric sets: applicability and limitations in protein folding and binding. *Protein Sci.* 4:2499–2509.

Le Grand, S. M., and K. M. Merz, Jr. 1993. Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. *J. Comp. Chem.* 14:349–352.

Merz, K. M. 1992. Analysis of a large data base of electrostatic potential derived atomic charges. *J. Comp. Chem.* 13:749–767.

Ooi, T., M. Oobatake, G. Nemethy, and H. A. Scheraga. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of protein. *Proc. Natl. Acad Sci. USA.* 84:3086–3090.

Pascual-Ahuir, J. L., E. Silla, and I. Tunon. 1994. GEPOL: an improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J. Comp. Chem.* 10:1127–1138.

Perrot, G., B. Cheng, K. D. Gibson, J. Vila, K. A. Palmer, A. Nayeem, B. Maigret, and H. A. Scheraga. 1992. MSEED: a program for rapid analytical determination of accessible surface areas and their derivatives. *J. Comp. Chem.* 13:1–11.

Ponnuswamy, P. K. 1993. Hydrophobic characteristics of folded proteins. *Prog. Biophys. Mol. Biol.* 59:57–103.

Rashin, A. A. 1990. Hydration phenomena, classical electrostatics, and boundary element method. *J. Phys. Chem.* 94:1725–1733.

Rashin, A. A., and K. Namboodiri. 1987. A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. *J. Phys. Chem.* 91:6003–6012.

Richards, F. M. 1977. Areas, volumes, packing, and protein structures. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

Sharp, K. A., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.

Simonson, T., and A. T. Brünger. 1994. Solvation free energies estimated from macroscopic continuum theory: an accurasy assessment. *J. Phys. Chem.* 98:4683–4694.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Spackman, M. A. 1996. Potential derived charges using a geodesic point selection scheme. *J. Comp. Chem.* 17:1–18.

Varshney, A., F. P. Brooks, Jr., and W. V. Wright. 1994. Computing smooth molecular surface. *IEEE Comput. Graph. Appl.* 14:19–25.

Vila, J., R. L. Williams, M. Vasquez, and H. A. Scheraga. 1991. Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitors. *Proteins Struct. Funct. Genet.* 10:199–218.

Vorobjev, Y. N., J. A. Grant, and H. A. Scheraga. 1992. A combined iterative and boundary element approach for solution of the nonlinear Poisson-Boltzmann equation. *J. Am. Chem. Soc.* 114:3189–3196.

Vorobjev, Y. N., and H. A. Scheraga. 1997. A fast adaptive multigrid boundary element method for macromolecular electrostatics in a solvent. *J. Comp. Chem.* 18:569–583.

Woods, R. J., M. Khalil, W. Pell, S. H. Moffat, and V. H. Smith, Jr. 1990. Derivation of net atomic charges from molecular electrostatic potentials. *J. Comp. Chem.* 11:297–310.

Zauhar, R. J. 1995. SMATR: a solvent-accessible triangulated surface generator for molecular graphics and boundary element applications. *J. Comp. Aided Mol. Des.* 9:149–159.

Zauhar, R. J., and R. S. Morgan. 1988. The rigorous computation of the molecular electric potential. *J. Comp. Chem.* 9:171–187.

Zauhar, R. J., and R. S. Morgan. 1990. Computing the electric potential of biomolecules: application of a new method of molecular surface triangulation. *J. Comp. Chem.* 11:603–622.